Anomaly Analytics and Structural Assessment in Process Industries

Martin Atzmueller M.ATZMULLER@UVT.NL
Tilburg University (TiCC), Warandelaan 2, 5037 AB Tilburg, The Netherlands

David Arnu DARNU@RAPIDMINER.COM
RapidMiner GmbH, Stockumer Straße 475, 44227 Dortmund, Germany

Andreas Schmidt SCHMIDT@CS.UNI-KASSEL.DE
University of Kassel (ITeG), Wilhelmshöher Allee 73, 34121 Kassel, Germany

Keywords: anomaly detection, exceptional model mining, sequential patterns, industry 4.0

Abstract

Detecting anomalous behavior can be of critical importance in an industrial application context: While modern production sites feature sophisticated alarm management systems, they mostly react to single events. In the context of process industries and heterogeneous data sources, we model sequential alarm data for anomaly detection and analysis, based on first-order Markov chain models. We outline hypothesis-driven and description-oriented modeling and provide an interactive dashboard for exploration and visualization.

1. Introduction

In many industrial areas, production facilities have reached a high level of automation: sensor readings are constantly analyzed and may trigger various forms of alarms. Then, the analysis of (exceptional) sequential patterns is an important task for obtaining insights into the process and for modelling predictive applications. The research project *Early detection and decision support for critical situations in production environments* (FEE) aims at detecting critical situations in production environments as early as possible and to support the facility operator in handling these situations, e. g., (Atzmueller et al., 2016a). Here, appropriate abstractions and analytics methods are necessary to adapt from a reactive to a proactive behavior.

Appearing in *Proceedings of Benelearn 2017*.

This paper summarizes the implementation of a comprehensive modeling and analytics approach for anomaly detection and analysis of heterogeneous data, as presented in (Atzmueller et al., 2017a).

2. Related Work

The investigation of sequential patterns and sequential trails are interesting and challenging tasks in data mining and network science, in particular in graph mining and social network analysis, e. g., (Atzmueller, 2014; Atzmueller, 2016b). In previous work (Atzmueller et al., 2016b), we have presented the DASHTrails approach that incorporates probability distributions for deriving transitions utilizing HypTrails (Singer et al., 2015). Based on that, the HypGraphs framework (Atzmueller et al., 2017b) provides a more general modeling approach. Using general weight-attributed network representations, we can infer transition matrices as graph interpretations.

Sequential pattern analysis has also been performed in the context of alarm management systems, where sequences are represented by the order of alarm notifications, e. g., (Folmer et al., 2014; Abele et al., 2013; Vogel-Heuser et al., 2015). In contrast to those approaches, we provide a systematic approach for the analysis of sequential transition matrices and its comparison relative to a set of hypotheses. Thus, similar to evidence networks in the context of social networks, e. g., (Mitzlaff et al., 2011) we model transitions assuming a certain interpretation of the data towards a sequential representation. Then, we can identify important influence factors.

Preprint of: Atzmueller M, Arnu D, Schmidt A (2017) Anomaly Analytics and Structural Assessment in Process Industries. Appearing in: Proceedings Benelearn 2017. Eindhoven, The Netherlands, 2017

3. Method

The detection and analysis of irregular or exceptional patterns, i.e., anomalies (Hawkins, 1980; Akoglu et al., 2015), in complex-structured heterogeneous data is a novel research area, e. g., for identifying new and/or emerging behavior, or for identifying detrimental or malicious activities. The former can be used for deriving new information and knowledge from the data, for identifying events in time or space, or for identifying interesting, important or exceptional groups.

In this paper, we focus on a combined detection and analysis approach utilizing heterogeneous data. That is, we include semi-structured, as well as structured data for enhancing the analysis. Furthermore, we also outline a description-oriented technique that does not only allow the detection of the anomalous patterns, but also its description using a given set of features. In particular, the concept of exceptional model mining, (Leman et al., 2008; Atzmueller, 2015; Duivesteijn et al., 2016) suitably enables such description-oriented approaches, adapting methods for the detection of interesting subgroups (that is, subgroup discovery) with more advanced target concepts for identifying exceptional (anomalous) groups. In our application context of an industrial production plants in an Industry 4.0 context, cf. (Vogel-Heuser et al., 2015; Folmer et al., 2017), we based our anomaly detection system on the analysis of the plant topology and alarm logs as well as on the similarity based analysis of metric sensor readings. The combined approach integrates both.

For sequential data, we formulate the “reference behavior” by collecting episodes of normal situations, which is typically observed for long running processes. Episodes of alarm sequences (formulated as hypotheses) can be compared to the normal situations in order to detect deviations, i.e., abnormal episodes. We map these sequences to transitions between functional units of an industrial plant. The results can also be used for diagnostics, by inspecting the transitions in detail. In summary, we utilize Bayesian inference on a first-order Markov chain model, see Figure 1. As an input, we provide a (data) matrix, containing the transitional information (frequencies) of transition between the respective states, according to the (observed) data. In addition, we utilize a set of hypotheses given by (row-normalized) stochastic matrices, modelling the given hypotheses. The estimation method outputs an evidence value, for each hypothesis, that can be used for ranking. Also, using the evidence values, we can compare the hypotheses in terms of their significance.

For modeling, we use the freely available RapidMiner (Mierswa et al., 2006) extension of HypGraphs,

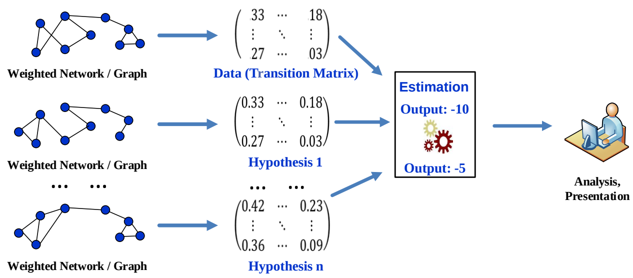

Figure 1. Overview on the modeling and analysis process.

that calculates the evidence values for different believe weights k and compares them directly with the given hypothesis and a random transition as a lower bound.

4. Process Model & Implementation

The first part of the analytical workflow is to build the transition network for training and testing the hypotheses. We build these hypotheses on real plant data and calculate the transition matrices for hourly time slots. In the same way, after further preprocessing (smoothing and down-sampling) we aggregate the corresponding raw sensor data. The calculated outlier score (Amer & Goldstein, 2012) is then presented, together with the evidence scores. A high outlier score indicates possible anomalous sensor readings and a low evidence score indicates deviating transition patterns in the alarm sequences. For further inspecting the outlier scores, we provide an additional dashboard. This shows the k highest outlier score for single sensor readings for a selected time segment and the associated sensor readings. Drilling-down from a high level of abstraction for a whole processing unit down to single sensor readings, a process engineer is then able to analyze possible critical situations in a convenient way.

For future work, we aim at extending the proposed approach by integrating the knowledge gained from a conceptual plant knowledge graph (Atzmueller et al., 2016a). We also plan to integrate the system into the Big data architecture proposed in (Klöpper et al., 2016), also considering further extensions on Big Data frameworks, e. g., (Meng et al., 2016; Carbone et al., 2015) and advanced assessment, exploration and explanation options, e. g., (Atzmueller et al., 2006; Atzmueller & Roth-Berghofer, 2010; Seipel et al., 2013) using advanced descriptive data analysis and modeling techniques, e. g., (Atzmueller, 2016a).

Acknowledgments

This work was partially funded by the BMBF project FEE under grant number 01IS14006.

References

Abele, L., Anic, M., Gutmann, T., Folmer, J., Kleinsteuber, M., & Vogel-Heuser, B. (2013). Combining Knowledge Modeling and Machine Learning for Alarm Root Cause Analysis. *MIM* (pp. 1843–1848).

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph Based Anomaly Detection and Description. *DMKD*, *29*, 626–688.

Amer, M., & Goldstein, M. (2012). Nearest-Neighbor and Clustering-based Anomaly Detection Algorithms for Rapidminer. *Proc. RCOMM* (pp. 1–12).

Atzmueller, M. (2014). Analyzing and Grounding Social Interaction in Online and Offline Networks. *Proc. ECML-PKDD* (pp. 485–488). Springer.

Atzmueller, M. (2015). Subgroup Discovery. *WIREs: Data Mining and Knowledge Discovery*, *5*, 35–49.

Atzmueller, M. (2016a). Detecting Community Patterns Capturing Exceptional Link Trails. *Proc. IEEE/ACM ASONAM.* IEEE Press.

Atzmueller, M. (2016b). Local Exceptionality Detection on Social Interaction Networks. *Proc. ECML-PKDD* (pp. 485–488). Springer.

Atzmueller, M., Arnu, D., & Schmidt, A. (2017a). Anomaly Detection and Structural Analysis in Industrial Production Environments. *Proc. International Data Science Conference.* Salzburg, Austria.

Atzmueller, M., Baumeister, J., & Puppe, F. (2006). Introspective Subgroup Analysis for Interactive Knowledge Refinement. *Proc. AAAI FLAIRS* (pp. 402–407). Palo Alto, CA, USA: AAAI Press.

Atzmueller, M., Kloepper, B., Mawla, H. A., Jäschke, B., Hollender, M., Graube, M., Arnu, D., Schmidt, A., Heinze, S., Schorer, L., Kroll, A., Stumme, G., & Urbas, L. (2016a). Big Data Analytics for Proactive Industrial Decision Support. *atp edition*, *58.*

Atzmueller, M., & Roth-Berghofer, T. (2010). The Mining and Analysis Continuum of Explaining Uncovered. *Proc. AI-2010.* London, UK: SGAI.

Atzmueller, M., Schmidt, A., & Kibanov, M. (2016b). DASHTrails: An Approach for Modeling and Analysis of Distribution-Adapted Sequential Hypotheses and Trails. *Proc. WWW 2016 (Companion).* ACM.

Atzmueller, M., Schmidt, A., Kloepper, B., & Arnu, D. (2017b). HypGraphs: An Approach for Analysis and Assessment of Graph-Based and Sequential Hypotheses. In *New Frontiers in Mining Complex Patterns*, LNAI. Springer.

Carbone, P., Ewen, S., Haridi, S., Katsifodimos, A., Markl, V., & Tzoumas, K. (2015). Apache Flink: Stream and Batch Processing in a Single Engine. *Data Engineering*, 28.

Duivesteijn, W., Feelders, A. J., & Knobbe, A. (2016). Exceptional Model Mining. *DMKD*, *30*, 47–98.

Folmer, J., Kirchen, I., Trunzer, E., Vogel-Heuser, B., Pötter, T., Graube, M., Heinze, S., Urbas, L., Atzmueller, M., & Arnu, D. (2017). Challenges for Big and Smart Data in Process Industries. *atp edition*, *01-02.*

Folmer, J., Schuricht, F., & Vogel-Heuser, B. (2014). Detection of Temporal Dependencies in Alarm Time Series of Industrial Plants. *Proc. 19th IFAC World Congress*, 24–29.

Hawkins, D. (1980). *Identification of Outliers.* London, UK: Chapman and Hall.

Klöpper, B., Dix, M., Schorer, L., Ampofo, A., Atzmueller, M., Arnu, D., & Klinkenberg, R. (2016). Defining Software Architectures for Big Data Enabled Operator Support Systems. *Proc. INDIN.*

Leman, D., Feelders, A., & Knobbe, A. (2008). Exceptional Model Mining. *Proc. ECML-PKDD* (pp. 1–16). Heidelberg, Germany: Springer.

Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D., Amde, M., Owen, S., et al. (2016). MLLib: Machine Learning in Apache Spark. *JMLR*, *17*, 1–7.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). Yale: Rapid prototyping for complex data mining tasks. *Proc. KDD* (pp. 935–940). New York, NY, USA: ACM.

Mitzlaff, F., Atzmueller, M., Benz, D., Hotho, A., & Stumme, G. (2011). Community Assessment using Evidence Networks. *Analysis of Social Media and Ubiquitous Data.* Heidelberg, Germany: Springer.

Seipel, D., Köhler, S., Neubeck, P., & Atzmueller, M. (2013). Mining Complex Event Patterns in Computer Networks. In *New Frontiers in Mining Complex Patterns*, LNAI. Springer.

Singer, P., Helic, D., Hotho, A., & Strohmaier, M. (2015). Hyptrails: A Bayesian Approach for Comparing Hypotheses about Human Trails. *Proc. WWW.* New York, NY, USA: ACM.

Vogel-Heuser, B., Schütz, D., & Folmer, J. (2015). Criteria-based Alarm Flood Pattern Recognition Using Historical Data from Automated Production Systems (aPS). *Mechatronics*, *31.*